VICTORIA UNIVERSITY OF WELLINGTON
Te Whare Wananga o te Upoko o te Ika a Maui

School of Mathematical and Computing Sciences
Computer Science

PO Box 600
Wellington
New Zealand

Tel: +64 4 471 5341, Fax: +64 4 495 5045
Email: Tech.Reports@mcs.vuw.ac.nz
Http://www.mcs.vuw.ac.nz/comp/Publications

Scholarly Web Sites: A General Method for Structuring Online Research Notebooks

Paul Martin

Technical Report CS-TR-CS-TR-98/7
August 4, 1998

Abstract

The researcher's notebook is a familiar tool in many disciplines and has several important uses: as a place to record detailed facts, as a place to store reminders of sources of useful information, and as a place to develop new ideas. In this paper we propose a general method for turning the research notebook into an online resource – what we call a 'Scholarly Web Site' – using a novel combination of directories and hypertext.

Keywords Hypertext, electronic notebooks, research methodology, WWW.

1 Introduction

The researcher's diary or notebook is a familiar tool in many disciplines. For our purposes, this paper-based friend has several important uses (Blaxter [2]). First, notebooks are used to look up specific facts, e.g. the page numbers of an article to be cited. Second, notebooks are used indirectly to find (or re-find) other material, e.g. the library's call number for an important text may be recorded in the notebook. Third, notebooks are used as a repository of information out of which new relationships and conclusions may be discovered. Fourth, notebooks – especially those in the form of loose-leaf folders or index cards – are used as a practice ground for marshaling and re-marshaling ideas until a logical disposition is found. Notebooks also have an important role in the verification of research (Zobel [6]) but this topic is beyond the scope of this paper.

The computer can extend the traditional paper-based notebook in two important ways, replacing the pen and paper to facilitate the production of a neat and tidy record, and providing a fast and effective search tool for locating information. The computer can also provide different 'views' of the material in the electronic notebook, as well as greatly speeding the re-marshaling of material.

2 A general method for structuring electronic notebooks

Our approach to producing online research notebooks uses standard WWW technology to structure hypertext nodes (Conklin [3]) into a 'Scholarly Web Site' (SWS). The researcher/author puts the contents of the notebook into a series of HTML files which are made available as Web pages using any conventional Web server. In our future work section we describe how these basic tools may be usefully augmented by some new tools.

The pages of a SWS are of two kinds: directory and content nodes. Since SWSs are not intended to exist in isolation but are intended to be part of the WWW, we also talk about 'external' nodes, meaning nodes not owned by the author of the SWS but directly linked to the SWS. The SWS has three kinds of link, directory, content and external, named according to the kind of destination node.

2.1 Directory nodes

The directory nodes are the main navigational device in a SWS and function in a similar way to the 'landmarks' described by Balasubramanian [1]. There are only a small number of directory nodes in any particular SWS and they are nodes which soon become familiar to the reader and help prevent her becoming 'lost in hyperspace' (Nielsen [4]). Directory nodes also function as familiar 'jumping off points' for the rest of the SWS.

Internally, a directory node is an ordered list of objects of the same type. For example, a SWS may have a directory of 'conference' type objects where each object represents knowledge acquired by the researcher about a particular conference. The type of an object determines the attributes of the object – a conference object might have 'title', 'acronym' and 'location' attributes. The intention is that the directory entry for a particular object will represent (directly or via links) *all* information known to the researcher about that object. Further, each object will have a rich network of directory or content links to every node in the SWS where it is mentioned, and each mention will have a (directory) link back to the object. A fragment of an example directory node for conferences showing three objects is given in Figure 1.

Directory links allow directory or content nodes to point to objects in directories and are annotated by parenthetically prefixing the hyperlink with the name of the directory in upper

Adaptive Hypertext and Hypermedia, Workshop on...

- AH '98, second, Pittsburgh, PA, June, in conjunction with Hypertext '98, {WWW}Home page
- AH '94, first, Cape Cod, Massachusetts, August, in conjunction with {CONFERENCE}UM '94

Education Multimedia and Hypermedia, World Conf. on...

- Ed-Media '99, Seattle, Washington, June, {WWW}Announcement

Figure 1: Directory node fragment

case. This shows that they are directory links (and not content or external links) and the type of the object at the link destination.

Figure 1 contains one directory link from the AH '94 conference object to the UM '94 conference object in the same directory. Figure 2 (below) contains two directory links into the 'ARTICLE' directory and one directory link into the 'PERSON' directory.

2.2 Content nodes

Content nodes contain largely unstructured information that doesn't fit naturally in any directory node. Examples of content nodes include descriptions and discussions of research topics, critiques of published articles, proposals for new work, and archives of work published in non Web-based forums. Figure 2 gives an example of a content node which explains the concept of a hypertext guided tour.

Guided tours The guided tour metaphor has been attributed to {ARTICLE}(Hammond 1987) but {ARTICLE}(Marshall 1989) also talks about guided tours, including the problems that authors have in selecting which nodes to include in each tour. For an online demonstration, see {PERSON}Watter's {WWW}Guided Tours Page. See also {hypertext concepts}trails, {hypertext concepts}paths, and {hypertext concepts}footprints.

Figure 2: Content node fragment

Content links are placed from content or directory nodes to content nodes and are annotated by parenthetically prefixing the hyperlink with some descriptive text in lower case to show that it is a content link and to explain the purpose of the link. Figure 2 contains three content links, all annotated with 'hypertext concepts'.

2.3 External nodes

External nodes are important since they often contain constantly updated or copyright information that can not be included by the researcher in the SWS. External links connect nodes within the SWS to nodes outside the SWS and are annotated with '{WWW}' to show that they are external and to warn the reader that the author accepts little responsibility for what is found at the destination.

2.4 An example SWS

The author has constructed, and has been using on a daily basis since 1997, a SWS for his research. This SWS contains 149 content nodes and just 6 directory nodes. This SWS has been a useful resource in its own right and a valuable testbed for development of the prototype tools described below. The cost of constructing the SWS has been somewhat greater than that of keeping a paper-based notebook, principally because of the need to maintain consistency in presentational format and because of the effort in creating and maintaining the hyperlinks. However, the SWS is considerably easier to browse than a (linear) paper notebook could ever be and the recent development of prototype tools for maintaining links has considerably reduced the effort of adding new information.

3 Evaluation of the SWS

Diverse kinds of information The traditional research notebook contains material of diverse kinds, for example raw data, lists, and draft papers. Being Web-based, the SWS can include inline any file format support by current browsers, plug-ins and helper applications. Further, as contemporary desktop applications are becoming increasingly 'Web aware' word processing documents, spreadsheets, and databases can be almost seamlessly integrated into the SWS.

Specific queries We noted earlier that researchers often need to look up specific facts. The SWS as currently proposed does not explicitly suppose the presence of a query-matching facility, although we note that most Web browsers allow the user to search for text in the current page and that a number of HTML authoring tools include a search engine for searching the pages in the author's Web site. Instead, our proposal anticipates that in most instances the SWS user will not try to formulate their query precisely but will navigate by browsing to find the information required.

Browsing Browsing is well supported in the SWS by means of the directory nodes, which act as 'jumping off' points, and rich interlinking (supported by tools we describe next), which ensures that every node is reachable by multiple, short paths. Further, our system for annotating links (also tool supported) helps users of the SWS interpret link function accurately and thus navigate efficiently.

Cognitive overhead and disorientation The consistency of the link annotations works to reduce 'cognitive overhead' (Wright [5]) while user disorientation is addressed by use of the directories as landmarks. Directories work well as landmarks because they are small in number, contain objects of the same type, and are largely independent of the application domain.

4 Future work

There are two key areas in which work on the SWS proposal still needs to be done. The first area involves evaluating the usability of the SWS with researchers in the field to see whether there are any criteria for effective online research notebooks that we have not anticipated.

The second area involves developing and evaluating software tools to automatically create and maintain the high degree of interlinking. Our prototype tools work by using headings in the text to identify and name objects in the directory nodes. To create directory links, the author need only plant a hint about the link required and the tools will substitute the correct HTML. For example, the tool will use string matching to deduce that the hint 'conf adaptive hypertext' refers to the 'Adaptive Hypertext and Hypermedia, Workshop on...' object in the 'CONFERENCE' directory.

We are also investigating automatic inference of new links – we know of several kinds of automatic inference that are possible, although we have yet to establish which are useful to the SWS author.

5 Conclusions

In this paper we have described the uses to which the traditional researcher's notebook is put and proposed a general method for structuring online versions of notebooks which we call Scholarly Web Sites. The unique characteristic of the SWS is its use of richly-linked directories as familiar 'landmarks' from which users browse the content.

References

[1] V. Balasubramanian. State of the art review on hypermedia issues and applications. Available from http://www.isg.sfu.ca/~duchier/misc/hypertext_review, December 1993.

[2] Loraine Blaxter, Christina Hughes, and Malcolm Tight. *How To Research.* Open University Press, Buckingham, United Kingdom, 1996.

[3] Jeff Conklin. Hypertext: An introduction and survey. *IEEE Computer*, 20:17–41, 1987.

[4] Jakob Nielsen. The art of navigating through hypertext. *Communications of the ACM*, 33(3):296–310, 1990.

[5] Patricia Wright. Cognitive overheads and prostheses: Some issues in evaluating hypertexts. In *Proceedings of Hypertext*, pages 1–12. ACM Press, New York, 1991.

[6] J. Zobel. Reliable research: Towards experimental standards for computer science. In *21st Australasian Computer Science Conference*, pages 217–229, Perth, Australia, February 1997.